Measuring Self—Regulation of Japanese EFL Learners in Relation to Vocabulary Acquisition

Measuring Self-Regulation of Japanese EFL Learners in Relation to Vocabulary Acquisition

Phillip Rowles

Abstract

The purpose of this study is to investigate fundamental measurement of self-regulation in regard to vocabulary acquisition. Participants in the study represent a sample of 72 Japanese EFL tertiary education level students from three anonymous institutions in Tokyo. In terms of receptive vocabulary breadth, they represent low, middle, and high ability groups. Four research questions address the following issues: a) differences in vocabulary acquisition, b) transferability of a self-regulation questionnaire, c) differences in self-regulation between groups, and d) differences between monolingual and bilingual versions of the self-regulation questionnaire. Rasch rating scale analysis is used to fundamentally measure and analyze the self-regulation questionnaire data. The study concludes with ideas for future EFL investigations into this timely, viable, but somewhat neglected area of interest in second language research--the self-regulation construct.

"Self-education is, I firmly believe, the only kind of education there is" (n.d.) is a quote attributed to Isaac Asimov, the author and bio-chemist. The focus of this paper is the measurement of another self-concept, self-regulation, especially related to vocabulary acquisition. In the process of writing this paper,

inspired by the prolific scribe Asimov, I endeavor to continue my self-education.

Psychological measurement is the focal point of Michell's (2005) *Measurement in Psychology*. A major dilemma outlined by Michell was that despite many psychologists' efforts to measure, the "presumption of successful psychological quantification is premature. One very disturbing sign is that many psychologists misunderstand what measurement is" (p. 4). Michell explains that "much of what passes for psychological measurement is based upon the counting of frequencies" (p. 5). However, Michell contends that one can move beyond this to measure intellectual abilities "only if it is known exactly how the probability of getting an item correct relates to the person's ability and the item' s difficulty" (p. 11). There are a number of theories related to this association, and Michell states the simplest relevant theory of these is the Rasch model. This paper will focus on the use of fundamental objective measurement and analysis as provided by the Rasch family of models. This paper is in answer to Zeidner, Boekaerts, and Pintrich's (2005) call for future studies to refine the measurement of self-regulation constructs in the *Handbook of Self-Regulation*.

The Rasch family of probabilistic models is based on the seminal work of Georg Rasch. Rasch's (1960) primary model proposed an original way to statistically estimate individuals' abilities and items' difficulties simultaneously based on a logistic function (Suen & French, 2003). Rasch's basic model dealt with dichotomously scored data. A Rasch model that could measure polytomously scored data, that is, like that used in Likert-scales, was developed from the work of Andrich (1978, cited in Bond & Fox, 2001) and fully realized by Wright and Masters (1982). Instead of investigating model-to-data fit as in classical test theory, the Rasch family of models investigates data-to-model fit (Bond & Fox, 2001). The Rasch model is "the only model to date that provides the tools for approximating objective reproducible additive measures in the

human sciences" (Bond & Fox, 2001, p. 8).

Testing of vocabulary is nothing new: the earliest large-scale assessment of human abilities in Western civilization was recorded in Judges (Judges 12:4-6, cited in Suen & French, 2003) in the Old Testament. A battle between two armies resulted in the Gileadites surrounding the Ephraimites. The Gileadites's dilemma was to identify the Ephraimites who looked the same as the Gileadites. One thing distinguished these two groups: the Ephraimites pronounced the word "Shibboleth" with an "s" sound, as opposed to the Gileadites, who pronounced it with a "sh" sound. In order to recognize absconding Ephraimites, suspicious captives were forced to say the word "shibboleth." Those who pronounced it with an "s" were killed. This high stakes instrument resulted in 42,000 people being killed (Suen and French, 2003).

The present study's measurement is not of such high stakes, in that nobody's life is in danger. Rather, the connection to the Judges story is through a very important component of language learning: vocabulary. Many may have perceived the Shibboleth test as a pronunciation test, however, broadly speaking, it was a vocabulary test. Learning vocabulary is wide reaching as there are eight types of information to acquire: phonological, orthographic, syntactic, morphological, pragmatic, articulatory, idiomatic, and semantic (Schreuder, 1987, cited in de Groot, 2006). If only the Ephraimites had acquired the phonological, pragmatic, and articulatory parts of vocabulary. The rest, as they say, would have been history. This is indeed an extreme example of the highest extrinsic motivation to pass a vocabulary test: the Ephraimities trying to save their own lives through displaying their vocabulary abilites.

My initial interest was in autonomy, and vocabulary learning strategies. However, I shifted my focus from strategies to *self-regulation* because of four main reasons: strategy definition difficulty, two explosions or booms in research,

and contemporary research.

The difficulty of defining strategies

The definition of learning strategies is fraught with ambiguity. Are learning strategies observable behaviors, inner mental operations, or both? (Dornyei, 2005; Ellis, 1994, cited in Tseng, Dornyei & Schmitt, 2006). A definition of learner strategies is problematic because of "the semantic-equivalence dilemma, with words like *strategy, operation, routine, process, procedure, action, tactic, technique, plan, and step*, being interchangeable in the literature" (Macaro, 2006, p. 324). Also, in learner strategy research, there has been a lack of theoretical rigor in general (Macaro, 2006).

"It needs little justification that a concept cannot be conceptualized as thought, belief, emotion, and observable behaviour at the same time; however this conceptual ambiguity is less of a problem in educational psychology, because, as we will see below, the term 'learning strategy' has virtually been abandoned for research purposes and has been maintained primarily for pedagogical discourse only" (Tseng et al., 2006, p. 80).

Research Explosion Number One

There was a boom in learning strategies research from 1980s to mid-1990s. However, there was a problem: conceptualizations of learning strategies as has been previously noted, were diverse.

Research Explosion Number Two

There was a boom in self-regulation research that started at the beginning of the 1990s. "It is not what learners do that makes them strategic learners but rather the fact that they put creative effort into trying to improve their own

learning. This is an important shift from focusing on the product – the actual techniques employed – to the self-regulatory process itself and the specific learner capacity underlying it. By the beginning of the 1990's, educational psychologists had gone through this transformation process and the study of self-regulation had come of age, causing a 'virtual explosion of work in this area' " (Zeidner et al., 2000, p. 750, cited in Tseng et al. 2006, p. 81).

Contemporary Research

Tseng et al. (2006) attempted to build on this second research explosion, and apply it in a specific (vocabulary acquisition) L2 context. They did this by proposing to transfer the self-regulation construct from educational psychology into the field of second language vocabulary acquisition. Tseng et al. (2006) acknowledged that in the past, strategic learning and learning strategies have been usually examined by the administration of self-report questionnaires. The *Motivated Strategies for Learning Questionnaire* (MSLQ) by Pintrich, Smith, Garcia, and McKeachie (1991, cited in Tseng et al. 2006) focused on the *quality* of items, so the scales could be assumed to be in a linear relationship with an underlying trait. Therefore, as the MSLQ scales were cumulative, the computing of MSLQ mean scale scores was perceived as psychometrically justifiable. The more commonly used *Strategy Inventory for Language Learning* (SILL) by Oxford (1990, cited in Tseng et al. 2006) focused on the quantity of items, so the scales were not assumed to be in a linear relationship with an underlying trait. Thus, as the SILL scales were not cumulative, the computing of SILL mean scale scores was not perceived as psychometrically justifiable. In other words, the SILL is problematic because a learner can achieve a high score if their response indicates that they use many strategies (quantity), regardless of how effectively they use them (quality).Despite these problems, the SILL has

been very influential on research. Even in the realm of vocabulary learning strategies, at least two major instruments have been influenced directly by the SILL taxonomy, including Stoffer's (1995, cited in Tseng et al.) Vocabulary Learning Strategies Inventory (VLSI), and Schmitt's (1997, cited in Tseng et al.) Vocabulary Learning Strategies (VLS) taxonomy.

Tseng et al.'s (2006) study proposed a new psychometrically valid instrument, the *Self-Regulating Capacity in Vocabulary Learning Scale* (SRCvoc) that examined self-regulation and vocabulary learning strategies. Instead of investigating behavioral habits (like the SILL did), the new instrument was inspired by the MSLQ's general declaration items. The theoretical construct used to base the SRCvoc was self-regulation strategies from educational psychology (Dornyei, 2001, cited in Tseng et al.), whichcombined action control strategy taxonomies (Kuhl, 1987, and Corno & Kanfer, 1993, in Tseng et al.). Five facets of control emerged: commitment, metacognitive, satiation, emotion, and environmental. The SRCvoc was developed over three phases: developing an item pool, running a pilot version, and giving the final instrument to the sample to be validated.

Tseng et al.'s (2006) study is problematic in a number of ways. The pilot study (phase 2) was given to a dissimilar sample compared to the main study. In the pilot, a sample was drawn from Taiwanese university students, while in the main study, the sample was drawn from Taiwanese high school students. The reason Tseng et al. focused onvocabulary acquisition was "the significance that we attach to the mastery of lexis in the second language acquisition process" (Tseng et al., p.79). Despite this significance, neither the pilot study nor the main study participants were tested for vocabulary proficiency measures. Assessment of the SRCvoc was presented in the form of Classical Testing Theory: reliability was checked by Cronbach Alpha internal consistency coefficients, and

evaluation was checked by confirmatory and exploratory factor analysis. There were reported satisfactory psychometric properties for the SRCvoc, and the model displayed good fit to the data. However, Tseng et al. (2006) stated this caveat "If we borrow the theoretical construct of self-regulation from education psychology, it still leaves the problem of how to operationalize it and measure it" (p. 82).

Purpose

The purpose of this study is to propose a response to Tseng et al.'s (2006) warning: attempting to operationalize and measure the theoretical construct of self-regulation. In order to achieve this , I will administer the SRCvoc to a sample of Japanese EFL university students. Before administration, I propose to test the sample for vocabulary acquisition levels using the most renowned instrument available, to eliminate guesswork as used in Tseng et al.'s (2006) study. The vocabulary test is called the Vocabulary Levels Test (VLT) originally developed by Nation (1983). However, this study will use an amended VLT version by Schmitt (2000). Moving beyond the ordinally scored data, or classical test theory, used for analysis in Tseng et al. (2006), I plan to transform the ordinally scored data into intervally measured data by using the Rasch rating scale model (Wright & Masters, 1982), and then analyzing the data using the Rasch rating scale model. Fundamental or objective measurement of self-regulation is the focus of this present study. To these ends, the following four research questions guide this study:

1. What are the receptive vocabulary breadth levels of a sample of Japanese EFL tertiary education learners?
2. Is there transferability of the *Self-Regulating Capacity in Vocabulary Learning Scale* (SRCvoc) (Tseng et al., 2006) to a Japanese EFL

tertiary education context?

3. Are there differences between low, middle, and high receptive vocabulary breadth level groups in relation to self-regulation of vocabulary acquisition?
4. Are there differences in participants responses when allowing learners to use a dictionary on a monolingual (English) questionnaire, and later a bilingual (English/Japanese) questionnaire?

Methods

Participants

An initial sample of 140 mainly first year Japanese tertiary education students were investigated from three anonymous institutions in Tokyo, Japan. These 140 students were from five different classes. From the initial sample of 140 students, 72 students from three different classes (Group A, B, and C, or high, middle, and low, respectively) were selected for investigation on the basis of results from the VLT. The selection process is dealt with in more detail in the results section in answer to research question one.

Table 1 displays the descriptive statistics for Group A, B, and C, which are the focus of the rest of this study. Group A came from an elite 'Tokyo Six' university, Group B came from a middle-level university, and Group C came from a women's junior college.

Materials

Photocopies were prepared of Schmitt's (2000) Version 1 amended VLT. Only two levels, the 2000, and Academic Vocabulary levels (hereafter called the 2K and AV) were administered to the students. The students used

Table 1

Descriptive Statistics for Group A (high-level, n=22), Group B (middle-level, n=24), and Group C (low-level, n=26)

Group	Major	Gender		Age				Tertiary Year	
		Men	Women	18	19	20	21	1st	2nd
A	Sociology	11	11	2	10	10	–	22	–
B	English Communication	21	3	5	19	–	–	24	–
C	English	–	26	5	15	5	1	20	6
Total (A+B+C)		32	40	12	44	15	1	66	6

Note. $N = 72$ participants.

pencils to complete the VLT. There were 30-items for the 2K and AV levels respectively, combined to make a total of 60-items.The VLT was chosen for two main reasons: it is the closest our profession has to a standard and accepted vocabulary instrument, and the VLT had never been administered to the participants in this study before. However, despite obtaining the receptive vocabulary breadth estimate from the VLT, it is still only a rough estimate. This is because of the sensitivity of the VLT and its design. Although participants were encouraged not to guess, the VLT lends itself to allow guessing through a process of eliminating correct and incorrect answers. See Appendix B for a cluster example fromSchmitt's (2000) amended VLT Version 1. The 2K and AV levels are made up of 10 clusters respectively, with each cluster having three items.

Photocopies were prepared of Tseng et al.'s (2006) English monolingual

version of the SRCvoc questionnaire. The 20-items covered five-subscales of control: commitment, metacognitive, satiation, emotion, and environment. A six-point Likert scale was used covering polytomous responses of: strongly agree, agree, partly agree, slightly disagree, disagree, and strongly disagree (left to right).

An original English/Japanese bilingual version of the SRCvoc was created and photocopied. For each item there was a statement in English and a statement in Japanese. A six-point Likert scale was used, however in this version, the responses were strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree (left to right).

Procedures

Two levels (2K and AV) of Version 1 amended VLT (Schmitt, 2000) were photocopied and distributed to the 140 participants in class time. The participants completed the test using a pencil. The tests were subsequently graded by hand using a template for efficiency and accuracy. From these VLT grades, three classes were selected representing high, middle and low classes from the original five class sample.

One week later, the participants in the three class sub-sample were administered the English monolingual SRCvoc version (Tseng et al., 2006) in class time. For analysis the Likert scale responses were reverse coded before analysis. This was because the Likert scale was backwards, in that it ran strongly agree to strongly disagree (left to right). Two exceptions were made: Items 1 and 12 as they were negatively worded. Unfortunately, Tseng et al. did not mention about these two points in their study.

A Japanese translation was made from the original English monolingual version of the SRCvoc using a group of Japanese native speakers. This

translated version was checked by two Japanese native speaking experts who had both completed PhD's in Applied Linguistics in English speaking countries. Approximately one month after the monolingual questionnaire administartion, the participants were administered an original English/Japanese bilingual SRCvoc version in class time.

The VLT and monolingual SRCvoc were administered during December, 2006, whilethe bilingual SRCvoc was administered during January, 2007. The participants were informed about the purpose of the study, also, that they were free to decide whether to participate or not, and the results would not be reflected in their final grade. The VLT took about 15 minutes to complete, while the monolingual SRCvoc took about 20 minutes and the bilingual SRCvoc took about 10 minutes.

Results

The results answer the four research questions outlined previously. The results section has therefore been divided up into four main parts to answer these research questions.

1. *What are the receptive vocabulary breadth levels of a sample of Japanese EFL tertiary education learners?*

From Appendix A, low, middle and high-level receptive vocabulary breadth class groups were selected, as scored by the VLT. The low-level group was Class 4 (M = 26.33), the middle-level group was Class 3 (M = 45.60), and the high-level group was Class 2 (M = 53.79). Only these three class groups (low, middle, and high) will be examined further in the remainder of this study. For clarity, the high-level group will hereafter be called Group A, the middle-level group, Group B, and the low-level group, Group C.

After examining the descriptive statistics in Appendix A, *z*-statistics were subsequently calculated. The z-skewness (skewness statistic divided by the skewness standard error) for Group A, Group B, and Group C scores was -1.98, -5.67, and 0.63, respectively. The *z*-kurtosis (kurtosis statistic divided by the kurtosis standard error) for Group A, Group B, and Group C scores was 0.80, 9.14, and -0.63, respectively. The *z*-skewness value for Group A (-1.98), and Group B (-5.67) fell outside the ± 1.96 criterion for acceptability of univariate normality for skewness. The *z*-kurtosis value for Group B (9.14) fell outside the ± 1.96 criterion for acceptability of univariate normality for kurtosis. However, as these values came from a relatively large total sample (N = 140), we can be assured that statistically significant deviations from normality will rarely make a substantive difference to an analysis. Therefore, despite these z-statistic values falling outside the ± 1.96 criterion, we can indicate that the dependent variable was normally distributed.

2. *Is there transferability of the Self-Regulating Capacity in Vocabulary Learning Scale (SRCvoc) (Tseng et al., 2006) to a Japanese EFL tertiary education context?*

Using Rasch rating scale analysis, the items of both the English monolingual (*M*) and English/Japanese bilingual (*B*) SRCvoc versions were examined. Before examination, the raw ordinally scored data was transformed into intervally measured data, specified in logits (or log odds ratios). The fit statistics of the logits were carefully examined. It was determined that weighted mean square residuals, or Infit Mean Square (MNSQ) values of between 0.6 and 1.4 (Wright & Linacre, 1994) indicated acceptable fit. Two items (1 and 12) in the *M* SRCvoc version and 3 items (16, 17, and 20) in the *B* SRCvoc version displayed misfit. See Table 2 for the displayed results. However, these items were retained in the analyses. For the monolingual questionnaire, Items 1 and

Table 2

Misfitting Items by Measure, Standard Error, and Infit Values from English Monolingual (M) and English/Japanese Bilingual (B) SRCvoc Versions

	M			*B*		
Item #	Measure	SE	Infit *MNSQ*	Measure	Infit SE	Infit *MNSQ*
1	1.01	0.12	1.97*	-0.67	0.12	1.14
12	0.15	0.12	1.62*	-0.20	0.12	1.23
16	0.07	0.12	0.66	0.31	0.12	0.56*
17	-0.14	0.12	1.17	-1.25	0.14	2.06*
20	-1.08	0.14	1.29	-1.10	0.13	1.61*
	Separation		Reliability	Separation		Reliability
Item	2.74		0.88	4.21		0.95
Person	2.22		0.83	2.34		0.85

Note. * indicates Infit Mean Square (*MNSQ*) value fell outside the productive measurement range of 0.6 to 1.4 (Wright & Linacre, 1994).

12 were negatively worded, and thus were probably misfit items. The reasons behind the bilingual questionnaire's Items 16, 17, and 20 misfit, may have something to do with the translations.

An indication of the SRCvoc questionnaire's sensitivity of participants self-regulation in vocabulary acquisition is provided by person strata (Weaver, 2005). Person strata indicates how many statistically different levels that the questionnaire can differentiate. Person strata is calculated by the equation: $(4G_p + 1)/3$. In Table 2, for the monolingualquestionnaire, person separation was 2.22 (r = .85) and item separation was 2.74 (r = .88). Therefore, for the monolingual questionnaire, person strata is (4*2.22 + 1)/3 = 3.29, which indicates about three statistically different groups of participants. For the bilingual questionnaire, using the results of Table 2, person separation was 2.34 (r = 0.85) and item

separation was 4.21 (r = 0.95). Therefore, for the bilingual questionnaire, person strata is (4*2.34 + 1)/3 = 3.45, which indicates about three statistically different groups of participants. Now we will examine the three statistically different groups.

3. *Are there differences between low, middle, and high receptive vocabulary breadth level groups in relation to self-regulation of vocabulary acquisition?*

The differences between the low, middle, and high groups for the English monolingual SRCvoc version are displayed in Appendix C. The items are listed according to overall measure (n=72). Therefore, overall for the monolingual questionnaire, Item 1 was the most difficult to endorse, while Item 20 was the easiest to endorse.

The differences between the low, middle, and high groups for the English/ Japanese bilingual SRCvoc version are displayed in Appendix D. The items are listed according to overall measure (n=72). Therefore, overall for the bilingual questionnaire, Item 7 was the most difficult to endorse, while Item 17 was the easiest to endorse.

4. *Are there differences in participants responses when allowing learners to use a dictionary on a monolingual (English) questionnaire, and later a bilingual (English/Japanese) questionnaire?*

There are differences between responses to the monolingual and bilingual questionnaires. Table 3 gives an overall display of these differences according to the six-point Likert scale responses on both questionnaire versions.

Changes in measures for each of the individual items is more illustrative of specific differences between both versions. Appendix E displays the items ordered by changes in measure, with Item 7 indicating the largest positive measure change, and Item 1 indicating the largest negative measure change.

Table 3

Category Frequencies, Average Measures, Threshold, and Category Fit for Monolingual and Bilingual SRCvoc Questionnaires

Category Label	Observed Count	Average Measure	Threshold	Infit Mean Square	Category Name
Monolingual					
0	22	-0.27	None	1.16	Strongly disagree
1	160	-0.08	-2.29	1.13	Disagree
2	316	0.08	-0.68	0.86	Partly disagree
3	450	0.45	-0.05	0.93	Partly agree
4	368	0.84	0.85	0.96	Agree
5	123	1.33	2.18	1.02	Strongly agree
Bilingual					
0	55	-0.81	None	1.27	Strongly disagree
1	232	-0.59	-2.16	0.91	Disagree
2	367	-0.13	-0.79	0.96	Partly disagree
3	382	0.31	0.03	0.88	Partly agree
4	302	0.77	0.76	0.93	Agree
5	101	1.37	2.16	1.07	Strongly agree

Attempting to explain some of these measurement changes involves a closer inspection of individual item wordings. Appendix F displays a wording analysis of items containing difficult English words or phrases and also items that were worded negatively. The word analysis was conducted by asking five classes of second year tertiary level students outside this study's sample to indicate problem words and phrases.

Discussion

Despite the VLT being a rough estimate, there was considerable variation between the three group means examined in his study. Although it may be rough, there were certain efficiency and ease of administration attributes to using the VLT 2K and AV levels to obtain a workable stratification between these three vocabulary ability groups. Instead of guessing what the levels might be, Tseng et al. (2006) could have strengthened their study by administering this 15 minute instrument to their sample.

The SRCvoc questionnaire is transferable to a Japanese EFL tertiary context. However, it may be better to run the Likert scale in an opposite direction to what Tseng et al. (2006) did. In other words, running from left to right, instead of having strongly agree, agree, partly agree, slightly disagree, disagree, and strongly disagree, (like Tseng et al.) having strongly disagree, disagree, slightly disagree, slightly agree, agree, and strongly disagree might be better. The use of 'partly agree' by Tseng et al. is an anomaly in this Likert scale as partly is not matched on the opposite side. Also some of the differences in thresholds seen in Table 3 did not fall within the recommended range of 1.4 to 5.0 logits (Linacre, 1999, cited in Weaver, 2005). The values that did not fall within this range for the monolingual questionnaire were the differences between thresholds 2 and 3 (0.63), thresholds 3 and 4 (0.90), and thresholds 4 and 5 (1.33). For the bilingual questionnaire, the differences not within the acceptable range were between thresholds 1 and 2 (1.37), thresholds 2 and 3 (0.76), and thresholds 3 and 4 (0.79). This suggests that a 5-point or a 4-point Likert scale might be better to use in the future.

It may also be better to have an English/Japanese bilingual version. Tseng et al. (2006) used a Chinese version of the questionnaire, however, they published an English version in the appendix of their study. With a bilingual

version, the participants may have a feeling of learning vocabulary as well as being asked their opinions. The use of negatively worded items, like Items 1 and 12, with a backwards running Likert scale could be especially confusing for students. As shown in Table 2, Items 1 and 12 (the negatively worded items) for the monolingual questionnaire showed misfit. Some of the items included very difficult words like 'procrastination,' and 'invigorate.' These difficult words might have to be rephrased into more easily understood items.

Person strata calculated for both the monolingual and bilingual questionnaires indicated that there were three statistically different person groups present for self-regulation of vocabulary acquisition. These macro differences can be examined more closely with the micro, or item specific, differences shown in Appendices C and D.

There were differences in the difficulty of responses to the monolingual and bilingual questionnaires. Appendix G shows the 'racked' data for both versions. Racking involves having the same people with two sets of data, for example, in this case, 72 persons with 40-items. This includes 20-items from the monolingual questionnaire and 20-items from the bilingual questionnaire (Wright, 1996, and Wright, 2003). In the case of these two questionnaire versions, Appendix G indicates that Item 7 in the bilingual questionnaire was the most difficult to endorse, while Item 20 in the monolingual questionnaire was the easiest to endorse.

Conclusions

Limitations of the Study

One of the major limitations of this study is my status as a novice Rasch model researcher. There are still so many points to learn about fundamental

measurement using the Rasch family of models. Another limitation may be the relatively small sample of size of only 72 participants. In subsequent studies I would like to survey a much larger sample of participants. Another area to work on would be to start the questionnaire development process from Phase 1, which might involve conducting focus group interviews to initiate the writing of original items for an original questionnaire, as noted by Tseng et al. (2006).

Implications

Using the SRCvoc questionnaire in other contexts is a worthwhile pursuit, Additionally, investigating other areas of self-regulated second language learning instead of vocabulary, as suggested by Tseng et al. (2006), is recommended.

Suggestions for Future Research

In the future, research should focus on developing new questionnaires into the self-regulation construct. As self-regulation is well covered in general education, but not so well covered in second language learning, there exists a gap to be filled. It is a viable research opportunity to be investigated further.

The process of administering and translating the SRCvoc questionnaire was a rewarding experience. In the future, I propose to create original questionnaire items from focus group interviews, pilot the study, amend, and administer the main questionnaire. With the responses from the main questionnaire, I propose to collect the responses in the form of item banking and gradually amend the questionnaire with each iteration.

Another area of interest is the use of online surveys. Some available online surveys

include SurveyMonkey (http://www.surveymonkey.com), QuestionPro

(http://www.questionpro.com), and Zoomerang (http://www.zoomerang.com). The most reasonably priced of the three is SurveyMonkey, so I plan to implement a SurveyMonkey questionnaire in the future.

Refinement of the measurement of self-regulation constructs answers the call outlined by Zeidner, Boekaerts, and Pintrich (2005) for future studies in the introduction to this paper. This paper represents a novice's tentative steps along the road to self-education, which was emphasized by the quotation by Asimov (n.d.) in the introduction to this paper. Self-regulation is an abundant research area, especially in the field of second language research where it also seems to be under-researched. Looking towards the future, I am extremely interested in pursuing self-regulated research in subsequent studies.

References

Andrich, D. (1978). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*, 581-594.

Asimov, I. (n.d.). Quotation. Retrieved February 3, 2007, from http://en.wikiquote.org/wiki/Isaac_Asimov

Boekaerts, M., Pintrich, P., and Zeidner, M. (2005). Handbook of self-regulation. Burligton, MA: Elsevier.

Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, N.J.: Lawrence Erlbaum.

Corno, L. & Kanfer, R. (1993). The role of volition in learning and performance. *Review of Research in Education, 19*, 301-341.

De Goot, A. (2006). Effects of stimulus characteristics and background music on foreign language vocabulary learning and forgetting. *Language Learning,*

56, 463-506.

Dornyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Kuhl, J. (1987). Action Control: The Maintenance of Motivational States. In F. Halish and J. Kuhl (eds.) *Motivation, intention, and volition*. Berlin: Springer.

Linacre, J.M. (1999). Investigating rating scale utility. *Journal of Applied Measurement, 3*, 103-122.

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *The Modern Language Journal, 90*, 320-337.

Michell, J. (2005). *Measurement in psychology*. Cambridge: Cambridge.

Nation, I. S. P. (1983). *Teaching and learning vocabulary*. Victoria University of Wellington: English Language Institute.

Oxford, R. L. (1990). *Language Learning Strategies: What Every Teacher Should Know*. New York: Newbury House.

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, Michigan: The University of Michigan.

QuestionPro. Online survey. Retrieved January 26, 2007, from http://www.questionpro.com

Schmitt, N. (1997). Vocabulary Learning Strategies. In N., Schmit & M. McCarthy (1997) (Eds.) *Vocabulary: Description, Acquisition and Pedagogy*. UK: Cambridge.

Schmitt, N. (2000) *Vocabulary in language teaching. Cambridge*: Cambridge UniversityPress.

Schreuder, R. (1987). *Het mental lexicon* [The mental lexicon]. Inaugural address, University of Nijmegen, The Netherlands.

Stoffer, I. (1995). University foreign language students' choice of vocabulary learning strategies as related to individual difference variables. PhD dissertation. University of Alabama.

Suen, H. K. & French, J. L. (2003). A History of the Development of Psychological and Educational Testing. In C. R. Reynolds & R. W. Kamphaus. (Eds.). *Handbook of psychological and educational assessment of children.*

SurveyMonkey. Online survey. Retrieved January 26, 2007, from http://www.surveymonkey.com

Tseng, W., Dornyei, Z., & Schmitt, N. (2006). A New Approach to Assessing Strategic Learning: The Case of Self-Regulation in Vocabulary Acquisition. *Applied Linguistics, 27*, 78-102.

Weaver, C. (2005). Using the Rasch Model to Develop a Measure of Second Language Learners' Willingness to Communicate within a Language Classroom. *Journal of Applied Measurement, 6*, p. 396-415.

Wright, B.D. (1996). Time 1 to Time 2 Comparison. *Rasch Measurement Transactions, 10*, p.478-481. Retrieved February 25, 2007, from http://www.rasch.org/rmt/rmt101f.htm

Wright, B.D. (2003). Rack and Stack: Time 1 vs. Time 2. *Rasch Measurement Transactions, 17*, p. 905-906. Retrieved December 20, 2006, from http://rasch.org/rmt/rmt171a.html

Wright, B.D. & Linacre, J.M. (1994). Reasonable Mean-square Fit Values. *Rasch Measurement Transactions, 8*, 370. Chicao: MESA Press.

Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Zeidner, M., Boekaerts, M., & Pintrich, P. (2005, 2001). Self-Regulation: Directions and Challenges for Future Research, pp.749-768. In Boekarts, M., Pintrich, P., & Zeidner, M. (2005). *Handbook of self-regulation*. Burlington, MA: Elsevier.

Zoomerang. Online survey. Retrieved January 26, 2007, from http://www.zoomerang.com

Appendix A

Descriptive Statistics for Combined VLT Results in the 2000-word plus Academic Vocabulary Levels (2K-word + AV) (60-items) Divided into Classes

	Tertiary Institution				
	A		B	C	
	Class				
	1	2	3	4	5
M	50.88	53.79	45.60	26.33	35.94
95% Confidence Interval for *M* (Lower Bound)	49.26	52.41	42.98	23.34	30.41
95% Confidence Interval for *M* (Upper Bound)	52.50	55.18	48.23	29.33	41.47
SD	4.56	3.65	7.03	8.02	11.12
Skewness	-0.26	-0.85	-2.44	.27	.21
Skewness *SE*	.41	.43	.43	.43	.54
Kurtosis	-0.89	.68	7.59	-.52	-.14
Kurtosis *SE*	.80	.85	.83	.83	1.04
n	33	29	30	30	18

Note. $N = 140$.

Appendix B

VLT Version 1 Example Cluster (Schmitt, 2000)

1. business
2. clock ..part of a house
3. horse ..animal with four legs
4. pencil .. something used for writing
5. shoe
6. wall

Appendix C

Measures (and Standard Errors) in Logits for English Monolingual Questionnaire

Item Number	All (N=72)	MA (n=22)	MB (n=24)	MC (n=26)
1	1.01 (0.12)	0.60 (0.21)	0.75 (0.21)	1.95 (0.23)
11	0.36 (0.12)	0.33 (0.21)	0.69 (0.20)	0.04 (0.23)
18	0.30 (0.12)	0.51 (0.21)	0.32 (0.20)	0.10 (0.23)
15	0.22 (0.12)	0.42 (0.21)	0.19 (0.20)	0.04 (0.23)
12	0.15 (0.12)	0.07 (0.21)	-0.22 (0.21)	0.76 (0.22)
4	0.13 (0.12)	0.29 (0.21)	0.24 (0.20)	-0.18 (0.24)
6	0.10 (0.12)	-0.10 (0.21)	0.15 (0.20)	0.30 (0.23)
7	0.07 (0.12)	0.20 (0.21)	0.24 (0.20)	-0.29 (0.24)
16	0.07 (0.12)	0.12 (0.21)	0.24 (0.20)	-0.18 (0.24)
8	0.06 (0.12)	-0.18 (0.21)	0.24 (0.20)	0.15 (0.24)
2	0.02 (0.12)	0.16 (0.21)	-0.01 (0.20)	-0.12 (0.24)
19	-0.01 (0.12)	0.03 (0.21)	-0.06 (0.21)	-0.01 (0.23)
5	-0.11 (0.12)	-0.05 (0.21)	-0.14 (0.21)	-0.18 (0.24)
10	-0.14 (0.12)	-0.10 (0.21)	-0.40 (0.21)	0.10 (0.23)
17	-0.14 (0.12)	-0.18 (0.21)	0.03 (0.20)	-0.35 (0.24)
9	-0.17 (0.12)	-0.31 (0.21)	-0.06 (0.21)	-0.01 (0.23)
13	-0.19 (0.12)	0.42 (0.21)	-0.62 (0.22)	-0.52 (0.24)
14	-0.25 (0.12)	-0.54 (0.22)	-0.10 (0.21)	-0.07 (0.23)
3	-0.39 (0.13)	-0.68 (0.22)	-0.18 (0.21)	-0.35 (0.24)
20	-1.08 (0.14)	-1.00 (0.24)	-1.29 (0.24)	-1.03 (0.26)

Appendix D

Measures (and Standard Errors) in Logits for English/Japanese Bilingual Questionnaire

Item Number	All (N=72)	BA (n=22)	BB (n=24)	BC (n=26)
7	1.18 (0.13)	1.24 (0.26)	1.60 (0.24)	0.94 (0.20)
8	0.58 (0.12)	0.14 (0.23)	-0.07 (0.20)	0.23 (0.19)
11	0.45 (0.12)	0.30 (0.23)	0.71 (0.22)	0.38 (0.19)
13	0.45 (0.12)	1.05 (0.25)	0.48 (0.21)	0.08 (0.19)
5	0.35 (0.12)	0.41 (0.24)	0.35 (0.21)	0.34 (0.19)
4	0.33 (0.12)	0.08 (0.23)	0.27 (0.21)	0.56 (0.19)
16	0.31 (0.12)	0.30 (0.23)	0.44 (0.21)	0.23 (0.19)
6	0.23 (0.12)	0.08 (0.23)	0.45 (0.21)	0.15 (0.19)
2	0.18 (0.12)	0.19 (0.23)	-0.03 (0.20)	0.38 (0.19)
8	0.10 (0.12)	0.14 (0.23)	-0.07 (0.20)	0.23 (0.19)
9	0.10 (0.12)	-0.37 (0.24)	0.35 (0.21)	0.19 (0.19)
15	0.07 (0.12)	-0.03 (0.24)	0.10 (0.20)	0.12 (0.19)
19	-0.01 (0.12)	-0.20 (0.24)	0.10 (0.20)	0.00 (0.20)
10	-0.10 (0.12)	0.14 (0.23)	-0.27 (0.20)	-0.11 (0.20)
12	-0.20 (0.12)	0.25 (0.23)	-0.31 (0.20)	-0.31 (0.20)
14	-0.28 (0.12)	-0.20 (0.24)	-0.43 (0.20)	-0.23 (0.20)
1	-0.67 (0.12)	-0.79 (0.25)	-0.90 (0.21)	-0.61 (0.21)
3	-0.73 (0.12)	-0.60 (0.25)	-0.99 (0.21)	-0.65 (0.21)
20	-1.10 (0.13)	-1.47 (0.27)	-1.14 (0.22)	-0.93 (0.22)
17	-1.25 (0.14)	-1.71 (0.28)	-1.34 (0.23)	-0.98 (0.23)

Appendix E

Measures of English Monolingual (M) Questionnaire and English/Japanese Bilingual (B) Questionnaire Items, Listed According to Measurement Change

	Measure			
Item	*M*	*B*	Change	Item Statement
7	0.07	1.18	1.11	When learning vocabulary, I believe I can achieve my goals more quickly than expected.
13	-0.19	0.45	0.64	I believe I can overcome all the difficulties related to achieving my vocabulary learning goals.
5	-0.11	0.35	0.46	When learning vocabulary, I have special techniques to keep my concentration focused.
18	0.30	0.58	0.28	During the process of learning vocabulary, I am confident that I can overcome any sense of boredom.
9	-0.17	0.10	0.27	When learning vocabulary, I think my methods of controlling my concentration are effective.
16	0.07	0.31	0.24	When it comes to learning vocabulary, I think my methods of controlling procrastination are effective.
4	0.13	0.33	0.20	When learning vocabulary, I have special techniques to achieve my learning goals.
2	0.02	0.18	0.16	When I feel stressed about vocabulary learning, I know how to reduce this stress.
6	0.10	0.23	0.13	I feel satisfied with the methods I use to reduce the stress of vocabulary learning.
11	0.36	0.45	0.09	When it comes to learning vocabulary, I have my special techniques to prevent procrastination.

Appendix E (Continued)

Measures of English Monolingual (M) Questionnaire and English/Japanese Bilingual (B) Questionnaire Items, Listed According to Measurement Change

	Measure			
Item	*M*	*B*	Change	Item Statement
8	0.06	0.10	0.04	During the process of learning vocabulary, I feel satisfied with the ways I eliminate boredom.
10	-0.14	-0.10	0.04	When learning vocabulary, I persist until I reach the goals I set for myself.
19	-0.01	-0.01	0.00	When feeling bored with learning vocabulary, I know how to regulate my mood in order to invigorate the leaning process.
20	-1.08	-1.10	-0.02	When I study vocabulary, I look for a good learning environment.
14	-0.25	-0.28	-0.03	When learning vocabulary, I know how to arrange the environment to make learning more efficient.
15	0.22	0.07	-0.29	When I feel stressed about my vocabulary learning, I cope with the problem immediately.
3	-0.39	-0.73	-0.34	When I am studying vocabulary and the learning environment becomes unsuitable, I try to sort out the problem.
12	0.15	-0.20	-0.35	When I feel stressed about vocabulary learning, I simply want to give up.
17	-0.14	-1.25	-1.11	When learning vocabulary, I am aware that the learning environment matters.
1	1.01	0.67	-1.68	Once the novelty of learning vocabulary is gone, I easily become impatient with it.

Appendix F

Item Wording Analysis of English Monolingual Questionnaire According to Measurement Change with English/Japanese Bilingual Questionnaire Responses

Item	Measure Change		Item description	
	Up (+)	Down (–)	Negatively-Worded	Difficult Words or Phrases
7	1.11		–	–
13	0.64		–	overcome
5	0.46		–	concentration
18	0.28		–	confident, boredom
9	0.27		–	concentration
16	0.24		–	procrastination
4	0.20		–	–
2	0.16		–	–
6	0.13		–	overcome
11	0.09		–	procrastination
8	0.04		–	eliminate, boredom
10	0.04		–	persist, make for myself
19	0.00		–	regulate, mood, invigorate

Appendix F (Continued)

Item Wording Analysis of English Monolingual Questionnaire According to Measurement Change with English/Japanese Bilingual Questionnaire Responses

Item	Measure Change Up (+)	Measure Change Down (−)	Item description Negatively-Worded	Item description Difficult Words or Phrases
20		-0.02	–	–
14		-0.03	–	efficient
15		-0.29	–	cope with
3		-0.34	–	unsuitable, sort out
12		-0.35	Yes	–
17		-1.11	–	aware
1		-1.68	Yes	novelty, impatient

Note. Italicized difficult words or phrases indicate they were especially difficult.

Appendix G

Wright Map of Racked Data for English Monolingual (M Items) and English/Japanese Bilingual (B Items) 40 Questionnaire Items (M = 20 Items and B = 20 Items)

```
Persons MAP OF Items
        <more> | <rare>
             . |
               |
               |
70             +
               |
               |
             . |
               |
               |
             # T|
             #  |   B_7
60           #  +T
           .##  S|   M_1
           ##   |   B18
         .####  |S  B_4   B_5   B11   B13   B16
         .###  M|   B_2   B_6
       ########  |   B_8   B_9   B15   B19   M11   M18
50     ######    +M  B10   B12   M12   M15   M16   M_4
                     M_6   M_7   M_8
           ##   S|   B14   M19   M_2   M_5
           .##   |   M10   M13   M14   M17   M_9
             .   |S  B_1   B_3   M_3
             .   |
                 T|   B20
40           #   +T  B17
                 |   M20
                 |
                 |
                 |
                 |
30               +
        <less> | <frequ>
```

Note. $N = 72$. Each "#" represents 2 persons.